

DIALOG(R) File 347:JAPIO  
(c) 2003 JPO & JAPIO. All rts. reserv.

04629922     \*\*Image available\*\*  
CHARACTER RECOGNITION POSTPROCESSING SYSTEM

PUB. NO.:        06-301822 [ JP 6301822    A]  
PUBLISHED:      October 28, 1994 (19941028)  
INVENTOR(s):    SASE SHINJI  
APPLICANT(s):   NEC CORP [000423] (A Japanese Company or Corporation), JP  
                  (Japan)  
APPL. NO.:      05-088886 [JP 9388886]  
FILED:          April 15, 1993 (19930415)  
INTL CLASS:     [5] G06K-009/72  
JAPIO CLASS:    45.3 (INFORMATION PROCESSING -- Input Output Units)  
JOURNAL:        Section:    , Section No. FFFFFFFF, Vol. 94, No. 10, Pg. FFFFFFFF,  
                  FF, FFFF (FFFFFFFFF)

#### ABSTRACT

PURPOSE: To perform stable character recognition postprocessing of less contradictions even at the time of simultaneously handling a solid written character string in the Japanese language, a character string of European and American words, and a character string including them together in the post-processing of character recognition.

CONSTITUTION: A collation range extracting means 11 generates a collation range table 21 from a character recognition result 20 based on blank information and punctuation mark/parenthesis information. Information indicating whether a punctuation just precedes each word or not and information indicating whether a punctuation just follows the word or not are described in a dictionary 22. A collation propriety discriminating means 13 compares these dictionary information with the collation range table 21 to preliminarily check the validity of the read-out dictionary.



DIALOG(R) File 345:Inpadoc/Fam.& Legal Stat  
(c) 2003 EPO. All rts. reserv.

12096100

Basic Patent (No,Kind,Date): JP 6301822 A2 941028 <No. of Patents: 001>

Patent Family:

Patent No	Kind	Date	Applic No	Kind	Date
JP 6301822	A2	941028	JP 9388886	A	930415 (BASIC)

Priority Data (No,Kind,Date):

JP 9388886 A 930415

PATENT FAMILY:

JAPAN (JP)

Patent (No,Kind,Date): JP 6301822 A2 941028

CHARACTER RECOGNITION POSTPROCESSING SYSTEM (English)

Patent Assignee: NIPPON ELECTRIC CO

Author (Inventor): SASE SHINJI

Priority (No,Kind,Date): JP 9388886 A 930415

Applic (No,Kind,Date): JP 9388886 A 930415

IPC: \* G06K-009/72

Language of Document: Japanese



(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平6-301822

(43)公開日 平成6年(1994)10月28日

(51)IntCl.<sup>5</sup>

G 0 6 K 9/72

識別記号

庁内整理番号

9289-5L

F I

技術表示箇所

審査請求 有 請求項の数 2 O L (全 4 頁)

(21)出願番号 : 特願平5-88886

(22)出願日 平成5年(1993)4月15日

(71)出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72)発明者 佐瀬 慎治

東京都港区芝五丁目7番1号日本電気株式  
会社内

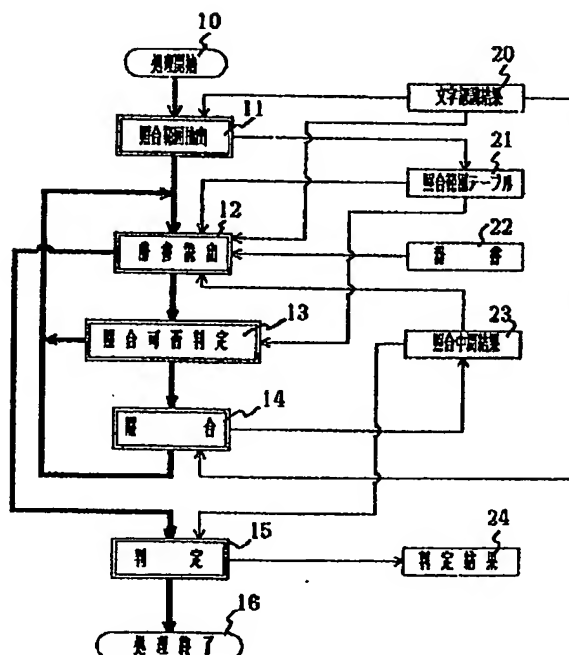
(74)代理人 弁理士 熊谷 雄太郎

(54)【発明の名称】 文字認識後処理方式

(57)【要約】

【目的】 文字認識の後処理において、日本語のべた書き文字列と欧米単語による文字列とこれらが混在する文字列を同時に取扱う場合でも、安定した矛盾の少ない文字認識後処理を行う。

【構成】 照合範囲抽出11は、文字認識結果20より、空白情報・句読点/括弧情報をもとに、照合範囲テーブル21を作成する。辞書22には、登録単語毎にその単語の直前が区切目であるという情報、その直後が区切目であるという情報が記述されている。照合可否判定13で、この辞書情報と照合範囲テーブル21とを比較し、読出した辞書の妥当性を事前にチェックする。



## 【特許請求の範囲】

【請求項1】 文字の並びに関する情報を記載した辞書をもとに、文字認識の結果を確認／補正する文字認識後処理方式において、空白等の情報をもとに入力文字列の照合範囲を定める照合範囲抽出手段と、文字認識結果およびそれまでの照合結果に応じて照合する情報を辞書より選択して読出す辞書読出手段と、読出した情報と情報毎に付帯された区切目情報をもとに照合可否を判定する照合可否判定手段と、照合可となる情報に対して文字認識結果との類似度を求める照合手段と、すべての照合結果より入力文字列に対する後処理結果を判定する判定手段とを有することを特徴とする文字認識後処理方式。

【請求項2】 前記照合範囲抽出手段は、文字認識結果を基にして、まず空白文字を抽出して照合範囲テーブルの該当位置を“1”にセットし、次に句読点、括弧等の区切位置を抽出して該当位置を“2”にセットし、その他の前記照合範囲テーブルの位置を“0”にセットし、前記照合可否判定手段は、辞書の単語に関する情報を基にして、前区切フラッグか後区切フラッグが“1”の場合には前記照合範囲テーブルの該当位置を参照し、該該当位置が“0”でなければ照合可能と判定することを更に特徴とする請求項1に記載の文字認識後処理方式。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、文字認識装置に関し、特に、読取結果を確認／補正する文字認識の後処理方式に関する。

## 【0002】

【従来の技術】 文字認識の後処理は、文字認識の不完全さを補助する手段として、広く活用されている。以前は、照合対象が単語単位であったものが、現在では文字列を照合単位とするまでに至っている。

【0003】 入力文字列の性質としては、欧米では単語単位で区切って書かれるのに対して、日本語では単語間に区切目を挿入せずに続けて書く（以後べた書きと称す）のが慣例となっている。

【0004】 このために、欧米文字列では、まず照合範囲を空白情報等で区切り、その間を一単語として照合を行う方法が一般的となっている。一方、日本語文はべた書きであるので区切目情報間を一単語とは想定せずに照合する方法をとっている。

## 【0005】

【発明が解決しようとする課題】 現状の日本語には英単語が併用されることが多々ある。また、日本語においても出現位置が文字列の区切目に特定できる場合がある。

【0006】 従来の欧米文方式の文字認識後処理方式では日本語のべた書きに対処できないという課題があり、日本語べた書対応の文字認識後処理方式では、欧米の単語が区切目情報なしに連続して抽出されたり、日本語においてもべた書の途中で、一続きの意味が終了したり、

途中から開始したりするという課題があった。

【0007】 一例として、べた書対応の文字認識後処理を「NOWHERE」という入力文字に対して実施すると、「NOWHERE」という一語単語と「NOW」「HERE」という二語単語が同じ確からしさで抽出されてしまうという課題がある。

【0008】 本発明は従来の上記実情に鑑みてなされたものであり、従って本発明の目的は、従来の技術に内在する上記諸課題を解決することを可能とした新規な文字認識後処理方式を提供することにある。

## 【0009】

【課題を解決するための手段】 上記目的を達成する為に、本発明に係る文字認識後処理方式は、空白等の区切目情報をもとに入力文字列の照合範囲を定める照合範囲抽出手段と、文字認識結果およびそれまでの照合結果に応じて照合する単語を辞書より選択して読出す辞書読出手段と、読出した単語と単語毎に付帯された区切目情報をもとに照合可否を判定する照合可否判定手段と、照合可となる単語に対して文字認識結果との類似度を求める照合手段と、すべての照合結果より入力文字列に対する後処理結果を判定する判定手段とを具備して構成される。

## 【0010】

【実施例】 次に本発明をその好ましい一実施例について、図面を参照して具体的に説明する。

【0011】 図1は本発明の一実施例を示すブロック構成図であり、処理の全体の流れを示す。

【0012】 図1を参照するに、処理開始10により、照合範囲抽出11が起動され、文字認識結果20をもとに、照合範囲テーブル21を作成する。次に辞書読出12で文字認識結果20と照合範囲テーブル21と照合中間結果23をもとに辞書22より照合に必要な情報を読み出す。

【0013】 必要な情報を読出した場合には照合可否判定13に、該当する情報がなくなった場合には判定15に処理をうつす。照合可否判定13は、照合範囲テーブル21を利用して、読出した情報の区切目情報の妥当性をチェックする。区切目情報が妥当な場合には、照合14に処理をうつし、妥当でない場合には辞書読出12に戻る。

【0014】 照合14は、辞書22から読出した情報と文字認識結果20の類似性を確認し、その結果を照合位置等と共に照合中間結果23に格納し、辞書読出12に戻る。

【0015】 判定15は、照合中間結果23より文字列全体の判定結果24を作成し、処理を終了する。

【0016】 本処理は、処理を実行する中央演算処理装置（CPU）と処理プログラム10～16と各種データ20～24を格納する記憶媒体（RAMとハードディスク、フロッピーディスク等）により構成することができ

る。

【0017】以下に処理11～13について詳細に説明する。なお処理14、15に関しては周知の技術で実現可能であるために、詳細な説明は省略する。

【0018】図2は文字認識結果20の一例を示す図であり、本実施例では文字認識結果として各文字位置に3つの候補が出力されている。照合範囲抽出処理11では、文字認識結果20を基にして、まず空白文字を抽出し、照合範囲テーブル21の該当位置を“1”にセットし、次に句読点、括弧等の区切位置を文字認識結果20より抽出し、照合範囲テーブル21の該当位置を“2”にセットする。その他の照合範囲テーブル21の位置は“0”とされる。本照合範囲抽出処理を図2の文字認識結果20に対して施した結果を図3に示す。

【0019】辞書読出処理12では、辞書照合範囲テーブル21の空白間の長さをもとに単語の長さ上限値を設定し、照合中間結果23をもとに単語の接続条件を設定し、文字認識結果20の候補文字を検索キー文字として、条件をみたす単語を検索して読み出す。辞書22には単語とその単語に関する情報が図4のような形式で格納されている。接続条件は図4の接続情報を、単語の長さは図4の単語長を、キー文字検索は図4の検索情報をそれぞれ利用する。

【0020】照合可否判定処理13では、図4のフラッグエリアが使用される。図5はフラッグエリアの中身を取出したものである。図の前区切は必ずその単語の直前に区切文字がある場合に“1”、そうでない場合には“0”が、後区切の場合には直後に区切文字がある場合には“1”が、そうでない場合には“0”がそれぞれセットされている。

【0021】例えば、英単語では多くの場合前区切/後区切共に“1”が、住所辞書の場合には日本住所の都道府県名には前区切に“1”が後区切に“0”がセットされている。

【0022】照合可否判定13は、辞書22の単語に関

するこれらの情報を読出し、前区切フラッグか後区切フラッグが“1”の場合には、照合範囲テーブル21の該当位置を参照し、テーブル21の該当位置が“0”でなければ照合可能と判定する。例えば、英単語「NOWHERE」は照合可能とされるが、「NOW」は照合不可能とされる。

【0023】

【発明の効果】以上説明したように、本発明によれば、単語毎に区切文字の情報を利用して、区切文字の必要の有無を確認しながら照合を行う構造を有しているため、日本語のべた書文字列と欧米文の混在する可能性のある文字列に対しても安定して矛盾のない文字認識照合を行う効果が得られる。

【図面の簡単な説明】

【図1】本発明の一実施例を示すブロック図である。

【図2】図1に示した文字認識結果20の一例を示す図である。

【図3】図1に示した照合範囲テーブル21の一例を示す図である。

【図4】図1に示した辞書22の一部分の例を示す図である。

【図5】図4に示した辞書のフラッグエリアの一部分の例を示す図である。

【符号の説明】

- 11…照合範囲抽出
- 12…辞書読出
- 13…照合可否判定
- 14…照合
- 15…判定
- 20…文字認識結果
- 21…照合範囲テーブル
- 22…辞書
- 23…照合中間結果
- 24…判定結果

【図2】

【図3】

【図5】

... F	NOWHERB	M ...	← 記入文字	... 01020000010...	
F	NOUHFRE	M	← 文字認識結果1位候補		
P	MOWMBBP	N	← 文字認識結果2位候補		
B	H.山N王PP	H	← 文字認識結果3位候補		

【図4】

単語ID	検索	情報	単語長	フラッグ	検索	情報	分類	文字1	文字2	...
------	----	----	-----	------	----	----	----	-----	-----	-----

【図1】

